Super-Resolved Surface Reconstruction From Multiple Images

Peter Cheeseman
RIACS
Bob Kanefsky
Robin Hanson
Sterling Software
John Stutz
NASA
Artificial Intelligence Research Branch
MS 269-2
NASA Ames Research Center
Moffett Field, CA 94035-1000

NASA Ames Research Center
Artificial Intelligence Research Branch

Technical Report FIA-93-02

February, 1993

Super-Resolved Surface Reconstruction From Multiple Images

Technical Report FIA-93-02

Peter Cheeseman
RIACS*

Bob Kanefsky
Sterling Software

Robin Hanson
Sterling Software

John Stutz
NASA
Artificial Intelligence Research Branch
NASA Ames Research Center, Mail Stop 269-2
Moffett Field, CA 94035, USA
Email: <last-name>@ptolemy.arc.nasa.gov
Phone: (415) 604-4946

January 29, 1993

Abstract

This paper describes a Bayesian method for combining information from multiple images of the same surface to construct a super-resolved surface model. We develop the theory and algorithms in detail for the 2-D surface reconstruction problem, and show the results on actual images. These results show dramatic improvements in resolution. The Bayesian approach uses neighbor correlation information as well as combining information from multiple images. The reconstructed surfaces have both significantly higher spatial resolution and grey scale resolution. We show how this theory can be extended to super-resolved 3-D surface reconstruction from multiple images, such as a series of frames from a moving TV camera. We also show that this approach can be applied to diffraction blurred images, offering the possibility of resolving images below the wavelength of light.

*Research Institute for Advanced Computer Science

1 Introduction

Consider the problem of how to extract as much information as possible from multiple images of the same scene, and of capturing this information in the form of a surface model at maximal resolution. This problem is important in many applications where maximal resolution is paramount, or where our goal is to discern fine surface detail. We here focus on space-based remote imaging.

We view this problem of surface reconstruction from multiple images as an example of an inverse problem—if we knew exactly the shape and reflectivity of the surface, the illumination conditions, the camera angle, etc., we could predict what the camera would see (pixels) to within the measurement accuracy. This is the image prediction problem addressed by computer graphics. Unfortunately, we have the inverse problem—we have the observed images (pixels) and must use this information to find the most probable surface that could have generated these images. Bayes theorem provides a formal solution to inverse problems, which we apply here to the surface reconstruction problem.

Because the reconstructed surface is only determined to within a certain maximum spatial resolution, we represent surfaces by a discrete uniform grid, at high spatial resolution, with the surface properties given at each grid point. For a planet like Mars, these surface properties include latitude, longitude, illumination, slope angle, slope direction, reflectivity at different wavelengths, etc. These are basically the properties that locate and characterize the grid point and describe how it could influence the image pixels. This surface grid is a reconstruction and is *not* what was actually observed. For this reason we call the surface grid values **mixels** (for Model **pixels**) to distinguish them from **pixels** which are the *observed* intensities. Unfortunately, in much of the vision literature, the word pixel is used interchangeably to refer to both inferred and observed values.

We are able to get super-resolved reconstructions from many images because each pixel of each image is a new sample of some patch on the observed surface. Two images generated with *exactly* the same alignment between the camera and the surface, and the same illumination conditions, might record the same information, resulting in no net gain of information. With slightly differing alignments however, the observed pixel values will be different. By relating these differences to locations on the surface, it is possible to reconstruct a model grid at a finer resolution than the observation pixilation. This technique for combining overlapping information is closely related to deconvolution (e.g. radar imaging) and computed tomography (e.g. CAT scan), and is explained in more detail in section 3.

We start by considering 2-D surface reconstruction. This is the best that can be achieved when the images are taken from essentially the same camera position, with the same relative sun angle, but with slight different registrations. This occurs with Landsat images, for example, where each location on Earth is imaged with only approximately constant location and orientation. The 2-D reconstruction gives the "reflectivity" of the surface, which is a combination of the effects of surface albedo, illumination conditions and ground slope. We develop this theory in detail in section 2, and show its application to actual images. This theory includes the use of prior knowledge in the form of neighbor correlations. In section 5, we outline how to extend this approach to full 3-D surface reconstruction, where images from different directions allow us to separate of the effects due to albedo from those due to ground slope.

2 2-D Surface Reconstruction

Our approach is based on Bayesian probability theory, and so uses a likelihood function, defined as the probability of the observed data given a model of how that data was generated. This model of the observation process is normally parameterized with respect to

anything that affects the process. For the surface reconstruction from multiple images problem, these observational parameters include surface illumination, atmospheric modulation, camera orientation, camera characteristics, optical distortions, and any data preprocessing. Computational considerations may require that some of these parameters be simplified or omitted, but doing so always entails some loss of precision. We have made several such simplifications in the work described in this section. The most important is that we model the surface as a 2-D plane lacking curvature and local relief. The second is the substitution of simple transformations (affine and quadratic) for the projective observation geometry and for any optical and electronic distortions of the camera system. A third lies in using a preprocessing step to deal with telemetry noise. For planetary images, we ignore atmospheric attenuation.

In our approach, we begin by constructing a likelihood function that gives the probability of each pixel, given the imaged surface and observation conditions. We take the likelihood of the entire image to be just the product of likelihoods of each pixel; the measurement error of a pixel is assumed to be (conditionally) independent of the value of its neighbors (i.e. there is no pixel cross-talk in the camera). This independence assumption is symbolically represented as:

$$\begin{aligned} &Pr[\text{all pixel values} \,|\, \text{observed params, surface model}] \\ &= \prod_p Pr[(\text{pixel(p)} = I_p \,|\, \text{observed params, surface model}] \end{aligned} \tag{1}$$

We assume that the probability of an observed pixel value is normally distributed, so that the likelihood of each pixel is given approximately by:

$$Pr[\text{pixel(p)} = I_p \,|\, \text{observed params, surface model}] \approx N[I_p \,|\, \hat{I}_p, \sigma]\Delta I_p \tag{2}$$

where subscript p refers to pixels, and $N[x \mid \mu, \sigma]$ is the standard normal (or Gaussian) distribution of x given a mean μ and standard deviation σ. ΔI_p is the difference between adjacent intensity values—i.e. the minimum observed grey-scale difference. This approximation is just the trapezoid approximation to the integral of a normal distribution over the interval from $I(p)$ to $I(p) + \Delta I_p$. The standard deviation σ of the observed pixels from their expected values is assumed to be the same for all pixels in all images. This deviation results from measurement error (especially quantization error) and model errors of various kinds (e.g. slight mis-registration, mixel value errors, etc.). If these many sources of error are largely independent, then (because of the central limit theorem) the resulting error distribution should be close to normal,

In Eqn. 2, the term $\hat{I}_p$ represents the expected intensity value for pixel p and is a complex function of the observation parameters and surface model. The parameters used in determining $\hat{I}_p$, as used in likelihood Eqn. 2 are:

1. **Registration Parameters:** These geometric parameters define how a pixel image maps onto the reconstructed mixel grid. Here, we use an affine transformation to define a 2-D (camera) to 2-D (mixel grid) function;

2. **Mixel Values:** This is the model of the reconstructed 2-D surface represented by a "reflectivity" or intensity value at each grid point (mixel);

3. **Point Spread Function (PSF):** This function defines how points on the surface (mixels) contribute to the observed pixels through the camera optics, as well as any distortion produced by camera readout;

4. **Camera Shading:** These parameters are necessary for cameras, such as Vidicon[1], with a nonuniform readout gain across the image plane. These parameters define a scaling factor that varies depending on where on the image plane a particular pixel falls.

The contribution of the above parameters to $\hat{I}_p$ is shown diagrammatically in Fig. 1. When all the mixels and parameters relating mixels to pixels are known, it is possible to

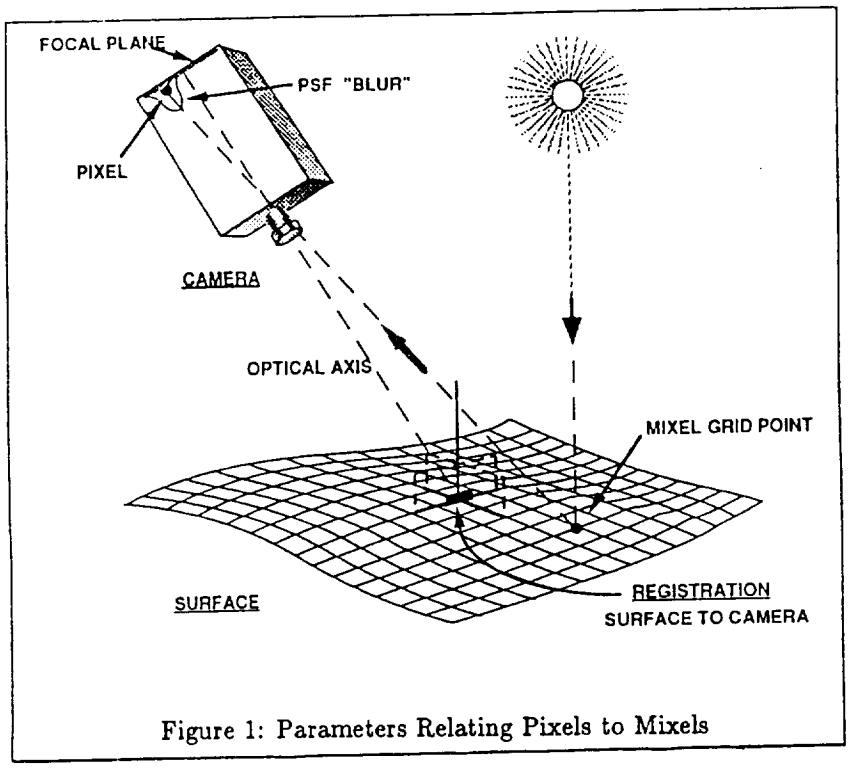

Figure 1: Parameters Relating Pixels to Mixels

calculate the expected value of a given pixel, $\hat{I}_p$, by summing the contribution of each mixel weighted by the PSF, as explained in the next section. This pixel prediction process is just a mathematical representation of the corresponding optical process shown in Fig. 1.

In a maximum likelihood (ML) approach, the goal is to find the set of parameter values that maximizes Eqn.(1)—in particular, the ML estimates of the mixels is a way of reconstructing an unknown surface from the multiple images. Note that finding the ML mixel values is a way of solving the inverse problem (i.e. model from data) given the likelihood (i.e. probability of data given model). When the resolution chosen for the mixel grid is *overdetermined* by the corresponding pixel values, the ML approach is a reasonable. The mixels are overdetermined by the pixels when there is no value for the mixels (except by chance) which can predict exactly all the pixel values. Unfortunately, the overdetermined situation means that the mixel grid is at a coarser spatial resolution than is otherwise achievable. If the ML approach is tried at too fine a resolution, the mixel values are underconstrained—i.e. there are many mixel grids that would predict the pixel values exactly, and there is no principled way of choosing among them.

The Bayesian approach used here is similar to the ML approach, but it uses additional (prior) knowledge in the form of expectations about correlations among neighboring mixels. This additional knowledge in the Bayesian maximum posterior (MP) estimate allows any

[1] A Vidicon camera is an obsolete electron beam readout camera, such as used in the Viking Orbiter images shown in this paper.

scale mixel grid. If too coarse a mixel grid is used (i.e. the mixels are overdetermined by the pixels), then the neighbor correlations have little effect, and the MP estimate is essentially the same as the ML estimate. However, if a very fine grid is used (i.e. the mixels are underdetermined), then the effect of the neighbor correlations competes with the fit to the data to give a reasonable compromise result that uses all the information. resolution is near the borderline between underdetermined, where the neighbor correlation the noise in the data. Details on finding the MP estimate are given in the next section.

3 MP Reconstruction

Given pixel data and a class of imaging models, we want to jointly estimate the mixel grid intensities together with other auxiliary model parameters. In a Bayesian approach one seeks a combination of all these parameters which has maximum posterior (MP) probability, which is the same (up to a normalization factor) as seeking a maximum joint probability:

$$\begin{aligned} \text{Joint Probability} &= \text{Likelihood} \times \text{Prior Probability} \\ P[\text{Mixels, Pixels, Params}] &= P[\text{Pixels} \,|\, \text{Mixels, Params}] \times P[\text{Mixels, Params}], \end{aligned}$$

where "Mixels" refers to the set of all mixel values, "Pixels" refers to the set of all pixels in all images, and "Params" refers to the auxiliary observational parameters (registration parameters, PSF, etc.) listed above.

Repeating Eqn. 2, the likelihood term is:

$$P[\text{Pixels} \,|\, \text{Mixels, Params}] = \prod_p N[I_p \,|\, \hat{I}_p, \sigma_p] \Delta I_p \tag{3}$$

We now specify the mean for each pixel $\hat{I}_p$ to be a linear combination of nearby mixels

$$\hat{I}_p = \sum_i \omega_{ip} m_i$$

where m_i is the intensity of the ith mixel and ω_{ip} is the mixel-pixel connection from the PSF and registration information, normalized to satisfy $\sum_i \omega_{ip} = 1$.

The prior probability term P[Mixels, Params] is the distinctly Bayesian contribution, and it embodies one's beliefs *before* seeing the data about the kinds of scenes or landscapes one might observe. Because the number of auxiliary model parameters is typically orders of magnitude fewer than the number of pixels or mixels, we can reasonably neglect the priors on these parameters, as they are highly over-determined. Instead, we use a conditional prior P[Mixels | Params] in the joint (instead of P[Mixels, Params]) and seek parameter values to maximize this new joint.

For simplicity we constrain the mixel prior to be a multivariate normal distribution over the set of mixel intensities, neglecting any source of prior information which can't be expressed in this form. Currently we use mean trends and neighbor correlations which are uniform across the mixel grid. Thus

$$P[\text{Mixels} \,|\, \text{Params}] = \mathcal{N}[m_i \,|\, \overline{m}, C_{ij}^{-1}] \prod_i dm_i \tag{4}$$

where subscripts i and j refers to mixels, and $\mathcal{N}[x_i \,|\, \mu_i, \Sigma_{ij}]$ denotes the joint multivariate normal distribution over some set of x_i given means μ_i and covariance matrix Σ_{ij}. Thus in Eqn. 4 $\overline{m_i}$ is the average mixel value, C_{ij} is the inverse of the mixel neighbor covariance matrix, and dm_i gives a differential volume element.

Note that while multivariate normals are usually described in terms of a covariance Σ_{ij}, we here define $C_{ij} = \Sigma_{ij}^{-1}$ to be an inverse correlation. With fractal landscapes, mixels may be highly correlated with other mixels a large distance away, making it impractical to directly use a covariance matrix. However, the inverse covariance matrix C_{ij} decays away more quickly with distance. Even more nicely behaved is the neighbor prediction matrix α_{ij}, defined from

$$C_{ij} = \frac{1}{s^2}(\delta_{ij} - 2\alpha_{ij} + \sum_k \alpha_{ik}\alpha_{jk})$$

where δ_{ij} is the delta function. This prediction matrix satisfies $\alpha_{ii} = 0$ and $\alpha_{ij} = \alpha_{ji}$, and gets its name from the fact that one can think of the landscape as generated by independent errors of the form $N[m_i \mid n_i, s]$, where the intensity m_i at each mixel i is best estimated by

$$n_i = \sum_j \alpha_{ij} m_j + \overline{m}(1 - \sum_j \alpha_{ij}),$$

a weighted average of intensities at neighboring mixels. In most of our tests we have used the isotropic values $\alpha_{ij} = 1/4$, when i and j are the N,S,E and West neighbors, and zero otherwise.

Our likelihood $P[\text{Pixels} \mid \text{Mixels, Params}] = \prod_p N[I_p \mid \sum_i \omega_{ip} m_i, \sigma_p]\Delta I_p$ estimates the I_p as a linear combination of the m_i. Therefore, the logarithm of this likelihood is bilinear in the m_i as well as in the usual I_p. Because of this, our likelihood can be re-written as proportional to a multivariate normal distribution in the m_i, since the logarithm of any multivariate normal distribution $\mathcal{N}[x_i \mid \mu_i, \Sigma_{ij}]$ is bilinear in all its x_i.

In terms of the m_i then, the total joint we seek to maximize is proportional to a product of two multivariate normal distributions—a likelihood and a prior term. But since the sum of two bilinear terms is itself bilinear, this total joint is itself a multivariate normal distribution

$$P[\text{Pixels, Mixels} \mid \text{Params}] = \mathcal{N}[m_i \mid m_i^*, A_{ij}^{-1}]$$

with total inverse covariance A_{ij} given by

$$A_{ij} = C_{ij} + \frac{1}{\sigma^2}\sum_p \omega_{ip}\omega_{jp},$$

and with a mean and peak m_i^* given by

$$\sum_j A_{ij}(m_j^* - \overline{m}) = \frac{1}{\sigma^2}\sum_p \omega_{ip}(I_p - \overline{m}). \tag{5}$$

Since it is a multivariate normal distribution, the value of this joint at its peak is maximized simply by maximizing the determinant of the matrix A_{ij}.

Thus we can find the maximum posterior mixel grid m_i^* given the auxiliary parameters by simply solving the above matrix equation, Eqn.5 and the determinant of A_{ij} can be used as a quality measure when choosing values for the auxiliary parameters. The method we use to actually compute the maximum posterior mixel grid m_i^* is discussed in the next section.

4 Reconstruction Algorithm

The previous section defined the basic MP mixel estimation equations to be solved; here we show how we solve these equations. The fundamental Catch-22 is that *if* we knew the true

values of the various parameters (PSF, registration etc.), we could solve the corresponding MP mixel equations. However, to estimate the necessary parameters to high accuracy, we would have to know the true mixel grid! Our way around this dilemma is to iteratively converge on a solution using our current best estimate of the mixels to re-estimate the necessary parameters, then to use these new parameter estimates to re-estimate a better mixel grid, and so on; hopefully converging on the MP mixel estimate.

4.1 Initial Registration Parameters

The registration parameters define the correspondence between points on the image plane and those on the modeled surface, and this is a trigonometric projection function that in principle varies for each point pair. This is basically a function of the imaging geometry; the imaging system's optics and its location and orientation relative to the surface. In the case of Vidicon cameras, there is an additional image distortion due to the read-out process that cannot easily be distinguished from geometric effects. The registration problem is to estimate the projection parameters for each image to the mixel grid that captures all these distortions.

In principle, the optimal MP registration parameters can be estimated using a prior distribution, and the likelihood of the pixel data given the PSF and the true image. Unfortunately, the true image is what we are trying to estimate, so we must initially use a surrogate. We randomly pick one to the pixel images as our surrogate, and interpolate its pixel values onto a grid to the desired resolution, as described below. Using this interpolated mixel grid as the reference, we search for accurate relative registration parameters (an affine transformation) that maps each image optimally onto the reference grid.

Instead of a MP estimate, we seek a simpler ML estimate, ignoring the parameter priors. This is rationalized on the basis of the large ratio of information (pixels) to the parameters that need be estimated. If we assume an independent Gaussian likelihood for each pixel relative to its projected value from the reference mixel grid, as in Eqn. 3, then finding the ML estimate of the registration parameters reduces to finding the registration with the smallest sum of squared pixel differences from their projected values (squared error). In other words, the optimal registration parameters for an image gives the minimum squared error when the projected mixel values are compared to the corresponding mixels through the PSF.

Optimal registration parameters were determined by the Simplex algorithm [1], which searches for a minima of the squared error by systematically varying the registration parameters, and then calculating the squared error for each registration. The algorithm stops when successive squared error values of the trial registrations are indistinguishable. Initial registration parameters for an image were obtained from its given nominal relation to the reference image. We found that unless the registration search starts relatively close to the true registration (good prior information), the search can get trapped in local minima. There are more efficient search algorithms than the Simplex algorithm, but they are not generally as robust. Note that previous methods for accurate relative image registration required locating "features" common to both images and finding a global mapping for all features to their counterparts in the other image [2]. The method described here uses *all* the information in both images, and this is part of the reason for the very high accuracy (better than 1/10th of a pixel) achieved by the method described here.

The Simplex search procedure presupposes a squared error that is a smooth function of the registration parameters. Unfortunately, this is not always true due to edge effects. If a variation of the registration parameters moves part of the pixel image across the model grid's edge, the likelihood undergoes large irregular variations, due to aliasing. We avoid this problem by defining a mixel grid border wide enough that no pixels ever reach the model's edge.

We have some evidence that a full affine transformation is not sufficient to capture all the distortions that occur in practice in a Vidicon image, but it is a very good approximation provided that not too large an area is being registered.

4.1.1 Interpolated Reference Grid

The initial reference image is expanded to the desired resolution and all the extra grid points between the initial pixels are interpolated with values that are a weighted average of the neighboring pixels. Initially, we tried a simple bilinear interpolation, which gives smooth interpolated values. Unfortunately, this interpolation did not work well for the following subtle reason. The interpolated values are a weighted average of the neighboring pixels, so they are better predictors of the mixel values than the pixels themselves because of statistical averaging. As a result, the registration search procedure occasionally became "stuck" when a trial registration happened to align with these interpolated values. An interpolation function was devised for us by Prof. Chris Wallace[2] that largely avoids this problem. It may be considered as an extension of bilinear interpolation that distributes the expected error as uniformly as possible across the grid. It was deduced from the requirements that, within the interpolated grid; the interpolation be everywhere continuous, the first derivative be everywhere continuous, the interpolation be linear with linear data, and the interpolation be equal for uniform data.

In one dimension, if we are given set of values Y_i at a set of regular grid points X_i, then our interpolated values $hatY(X)$ depends on data at the three nearest grid points:

$$\begin{aligned} hatY(X) &= Y_{i-1}\frac{1}{3}\left(1-\frac{1}{2}\sqrt{1-3\delta_x^2}-\frac{3}{2}\delta_x\right) \\ &\quad +Y_i\frac{1}{3}\left(1+\sqrt{1-3\delta_x^2}\right) \\ &\quad +Y_{i+1}\frac{1}{3}\left(1-\frac{1}{2}\sqrt{1-3\delta_x^2}+\frac{3}{2}\delta_x\right) \end{aligned}$$

as long as $\delta_x \equiv X - X_i$ satisfies $-1/2 \leq \delta_x \leq 1/2$. The multidimensional form is just the product of successive single dimension interpolations. The basic unit for interpolation is the unit point. Within this region, it uses the 3^n nearest grid points. On the region boundary, these factors shift from 3 to 2 for each lost degree of freedom.

4.2 PSF and Other Parameters

The point spread function (PSF) describes how the energy from an imaged point is distributed over the image plane. It is usually due to the optical system's diffraction and aberration pattern. With the scanning electron beam detector used in a Vidicon, the PSF can be extended to model the diffuse readout spot as well. Since the PSF is a function of the imaging system, it does not depend on the particular image. In practice, the PSF can vary across the image plane, and with time. We have not attempted to model this variance, working with an average PSF derived from the instrument's bench calibration [3].

"Shading" is the characteristic smooth variation in detector sensitivity across the image plane in Vidicon tubes. It's equivalent to the variation of individual cell sensitivities in array detectors. Shading must be corrected for in the likelihood model. The shading function can be learned from the data, given a rough idea of the registration. Since all images contain the same subregion under similar lighting and viewing angles, any systematic differences in their appearance must be due to shading. We assume the shading function is a low order 2-D

[2] Computer Science, Monash University, Australia

polynomial, and currently search for coefficients which make the subregions have the most similar mean intensities. These shading coefficients are used to preprocess the subimages for registration and restoration.

Defects in the optical system or on the image plane generate blemishes common to all images. A blemish map is used to identify and ignore suspect pixels. Also, in space-based remote sensing, there are potential errors in data transmission. Random or regular data blocks may lost, and there are bit errors in the telemetry stream. Lost pixels are mapped and ignored as blemishes on an individual image basis [4]. Telemetry noise is currently identified by deviation from neighbors, mapped, and ignored as blemishes.

4.3 Initial Composite

Having found good initial estimates of the basic parameters (PSF, registration parameters etc.), we next construct a composite mixel grid using information from *all* the pixel images. We construct the value of a composite mixel by calculating the "votes" from every pixel that could effect it, weighted through the PSF. These "votes" are accumulated to give a total mixel value

$$m_i = \frac{\sum_p \omega_{ip} I_p}{\sum_p \omega_{ip}}$$

for each mixel. Clearly, those pixels that are nearest the projected position of a mixel have the strongest vote for that mixel. For very narrow PSFs, the pixel-mixel "voting" is almost 1-to-1, but for very diffuse PSFs, each mixel value is the weighted combination of information from many pixels, leading to a "blurred" composite. In Fig. 2 (b),(e), the composite mixel grids under two different PSFs are shown—as can be seen, both composites are better than the original pixel images.

4.4 Iterative Improvement

Having obtained a reasonable starting composite mixel grid, we next sharpen it up, by iteratively converging it toward the MP estimate. We use a standard iterative method (Jacobi's method) to solve the essential matrix equation (5). Our resulting iterative mixel re-estimation formula is:

$$\Delta m_i = \alpha \frac{\frac{s^2}{\sigma^2} \sum_p \omega_{ip} (I_p - \hat{I}_p) - (m_i - n_i) + \sum_j \alpha_{ij} (m_j - n_j)}{\frac{s^2}{\sigma^2} \sum_p \omega_{ip}^2 + 1 + \sum_j \alpha_{ij}^2} \tag{6}$$

where α is set to obtain a converging, rather than diverging, iteration.

The results of applying this iterative formula to initial composite mixel grids is shown in Fig. 2 (c),(f); a noticeable sharpening of the composite is clear. When the mixel grid resolution is too coarse, the mixels are overdetermined by the pixels, so the MP mixel estimate is essentially the same as the ML estimate. In Eqn. (6), this means that the first (data) term in the numerator dominates the other two (mixel neighbor correlation) terms. When the mixel grid resolution is large enough (underconstrained by the pixels), the two terms in the numerator balance each other—i.e. the data term tries to force the mixels to exactly agree with the data, while the mixel neighbor term tries to make all mixels look like their neighbors ("smoothing"). It is the tension between these two effects that leads to plausible images, even when the mixels are underconstrained by the data, as is evident in Fig. 2 (a,b,c).

In Eqn. 6, all the necessary parameters (s, σ, and the registration, PSF, etc. parameters that go into ω_{ip}) are assumed known. Some of these parameters, such as the PSF, are

often well known ahead of time. Other parameters, such as the registration, can be initially estimated from an interpolated version of a single image. Since we find a much more probable mixel grid as a result of compositing and iteration, we can then re-estimate these parameters, and even repeat this convergence cycle. Fortunately, this re-estimation is not needed in practice more than twice. The reason for this is that parameters, such as the registration parameters, are typically estimated from thousands of pixels in the interpolated initial mixel grid, and so are already very accurate.

The deviation parameters σ and s are more difficult to estimate, as their most probable values can be many orders of magnitude different from what one might estimate from a composite. We initially intended to re-estimate these parameters *during* the iterative convergence cycle from the residual error in each new mixel grid. But being AI researchers and not experienced numerical programmers, we did not realize that in some cases this dynamic re-estimation would result in *diverging* from the true answer. So now when prior information is not enough to set these parameters, we must resort to an explicit search. We take a small but hopely representative patch of an image and seek parameters values which maximize our quality measure, the determinant of the matrix A_{ij}.

4.5 Results

We give results for a cartoon subject and for Viking Orbiter images of Mars [5]. The cartoon subject is a binary image. A dozen pixel images were generated, with a 4:1 pixel ratio (16:1 area ratio), using randomly chosen affine transforms and 10% Gaussian noise. These were composed with the known registrations, and then restored at a 1:4 mixel ratio. This system is both underconstrained and relatively noisy, yet the error between mixels and cartoon pixels has been significantly reduced, as shown in Fig. 4.5. Fig. 4.5 shows the true image from which the pixels images were extracted.

The Viking reconstruction uses a series of 24 Vidicon images of Mars. We extracted 128 x 128 pixel regions containing the same four prominent craters. These regions represent the same area to within a few pixels. The images were preprocessed using the techniques described in section 4.2. Vidicon blemishes and telemetry noise, were mapped and subsequently ignored, the shading response was modeled and used to correct the data, and image registration used the affine transforms. Restoration was done at a 1:4 mixel scale (1:16 area ratio), thus the restoration is slightly overconstrained. We leave it to the reader to judge the restoration's quality.

5 Extensions

In some applications, such as with Landsat and Voyager images, there are multiple images in different spectral bands. The 2-D reconstruction described above can be used on each spectral band separately, to get super-resolved surfaces for each band. However, this approach ignores the fact that the surface features are often very similar across bands, so even higher resolution is possible if these correlations are used.

In the above we discussed 2-D surface reconstruction, which is appropriate if all the images are from essentially the same direction under the same illumination conditions. However, for most of Voyager and Viking data, for example, there are many views of the same surface taken from different directions with different illumination. The theory described above can in principle be extended to handle this case as well. The essential idea is to find the MP 3-D surface model at a given resolution. This MP model predicts the observed images to within the noise level, and balances the fit to the data with prior knowledge of surface properties, such as continuity, smoothness, texture etc.

A possible approach to 3-D surface modeling is to represent the surface by a latitude-longitude grid, where each grid point is assigned an elevation, as well as other surface properties, such as albedo or ground-cover. Ground slope at a grid point can be estimated from the elevation of its neighbors; whether a grid point is in shadow can be deduced from the elevation of other grid points and the illumination direction; while illumination from secondary scattering can also be deduced from knowledge of the surface and illumination conditions. Given a 3-D surface model, illumination conditions, camera direction etc., it is possible to deduce what the camera will see. The Bayesian approach described above can be used to invert this process and find the most probable 3-D surface model that could have generated the images. This inversion process separates out the effects due to ground slope from those due to different albedos. This is possible because the effects of parallax vary independently of effects due to surface albedo. Note that representing the surface reflectivity by a single scalar (albedo) is an approximation that assumes Lambertian (cosine) scattering. Many real surfaces are not Lambertian, so using more complicated bi-directional reflectance parameters, including a specular reflectance component, would give a more accurate surface model.

The above 3-D surface model assumes that the surface itself does not change between successive images, but this is not true for dynamic situations, such as a TV camera observing moving people. While it is possible in principle to develop dynamic 3-D surface models so that super-resolution reconstructions can be performed for them, this much harder than the static case. However, such an extension would make it possible to get high definition TV from standard resolution cameras.

In all the above discussion, it was assumed that the multiple images captured by the camera were not significantly blurred by diffraction effects. When this is true, intensities contributed from different mixels are added to give the resulting pixel estimate (a linear model). If the phase of the signal from different mixels causes interference effects when they are combined (i.e. adds amplitudes not intensities) then the linear theory is not applicable. This situation applies to diffraction blurred microscope or telescope images, as well as various radar and sonar image problems, where the conventional wisdom is that it is not possible to break the Rayleigh limit—i.e. super-resolution is impossible. We believe that it *is* possible to extend the theory above to cover the diffraction case as well, so that in principle super-resolution is possible.

If the approach in this paper is changed from adding the intensities of different mixels to get a pixel estimate, to adding amplitudes instead; the resulting equations are again linear. Unfortunately, we cannot observe amplitudes, only amplitudes squared (i.e. intensities), so that the inversion problem in the diffraction case is inherently nonlinear. One way around this difficulty might be to do the calculations using amplitudes, and use the square root of the observed pixel intensities as the observational data. Because the square root has an ambiguous sign, prior knowledge, such as neighboring amplitudes probably have the same sign can be used to assign signs to the pseudo-observational amplitudes.

6 Relation to Previous Work

The research reported in this paper was mainly motivated by attempts to integrate information from Landsat images taken on different passes. The difficulty here is that such images did not exactly overlay each other, so pixel-to-pixel comparison is not possible. A standard approach to this problem is "rubber-sheeting", which attempts to fit one image grid to another (reference) grid by resampling the first image onto the reference grid. Reference grid points are mapped, through an appropriate transform, onto the new image, and new grid elements are computed by taking an area weighted average of the overlain image pixels.

The resulting resampled grid is perfectly aligned with the reference grid. The technique is extensively used to rectify and rotate Landsat and similar images to fit the geographical survey grid.

From the Bayesian perspective, the rubber-sheeting approach makes little sense, because the new averaged "pixels" are not actual observations, nor are they a surface model. In addition, the averaging process destroys information—it is impossible to recover the original image from the rubber-sheeted image. This loss of information makes pixel-by-pixel comparison very dubious. The super-resolved surface modeling described in this paper does allow the integration and comparison of information from many images through the accumulated super-resolution surface model.

A related approach to the Bayesian 3-D surface reconstruction described above is called "Shape from Shading" [6]. This approach integrates observed surface intensity gradients from a single image to give a 3-D elevation model of the generating surface. There is also a strong constraint due to surface continuity. This approach assumes a constant albedo and known illumination conditions. Shape from shading can be extended to multiple images [7], and the result is greater detail in the elevation map because each grid point contains information from multiple images. However, the constant albedo assumption is a strong limitation on the ability to extract information from multiple images.

A Bayesian approach very similar to ours is described in [8]. This approach does 3-D surface reconstruction using multiple images from different viewpoints, and a neighbor correlation prior with a Gaussian noise model. The representation of the surface is 2-D patches on a 3-D curved surface. Unlike our work, these authors assume smooth large scale objects that can be represented by large parameterized "surface patches". Because these patches are estimated from many pixels from many images, the parameters that describe them are accurately determined, and so the overall surface is accurately estimated. In our approach we achieve super-resolution, and there is no aggregation of surface mixels into large scale patches. Although our goals and assumptions are significantly different we use the same basic Bayesian approach.

7 Summary and Discussion

In this paper we have developed the Bayesian maximum posterior probability approach for super-resolved 2-D surface reconstruction from multiple images. We have also developed an algorithm for iteratively converging on the maximum probability estimate of the surface, and show results of applying this algorithm to a some images. Although we expected this approach to achieve super-resolution, we were surprised by how much spatial improvement was possible. In addition to improved spatial resolution, there is also improved grey-scale resolution in the reconstructed surface because of statistical averaging. The Bayesian approach can achieve a resolution higher than that for maximum likelihood, because it uses a neighbor correlation prior to pick out the smoothest image compatible with the data. This approach also is open to incorporating more sophisticated surface prior information, such as texture, edges, regions, etc.

Our 2-D approach can be extended to reconstructing super-resolved 3-D surface, but the surface models are much more complex than for 2-D. A possible application of this is to achieve high definition television by integrating the information between frames. It is also possible to extend this approach to "unblurring" diffraction blurred images, such as from microscopes or telescopes, although this extension requires difficult nonlinear mathematical extensions. This means that it is possible to resolve objects finer than the wavelength of the light used to observe them.

Acknowledgements

We gratefully acknowledge the fruitful contributions of Chris Wallace, Wray Buntine and Rich Kraft.

References

[1] W. Press, B. Flannery, S. Teukolsky, W. Vetterling,*Numerical Recipies in C*, Cambridge University Press, 1988.

[2] R. W. Gaskell.*Digital Identification of Cartographic Control Points*, Photogrammetric Engineering and Remote Sensing, Vol. 54, No. 6, Part 1, June 1988, pp. 723-727.

[3] M. Benesh, and T. Thorpe.*Viking Orbiter 1975 visual imaging subsystem calibration report*, JPL Document 611-125, Jet Propulsion Laboratory, Pasadena, Ca., 1976.

[4] E. Eliason et. al.*Adaptive box filters for removal of random noise from digital images*, Photogrammetric Engineering and Remote Sensing, 56, 453-456, 1990.

[5] M. Carr et. al.*Archive of Digital Images from NASA's Viking Orbiter 1 and 2 Missions*, Planetary Data System, National Space Science Data Center, CD-ROM: USA_NASA_PDS_VO_1003, Frames VO217S42-88.

[6] B.K.P. Horn and M.J. Brooks.*Shape from Shading*, MIT Press, Cambridge, Massachusetts, 1989

[7] J. Thomas, W. Kober, and F. Leberl.*Multiple Image SAR Shape-from-Shading*, Photogrammetric Engineering and Remote Sensing, Vol. 57, No. 1, Jan. 1991, pp. 51-59.

[8] Y. P. Hung and D. B. Cooper.*Maximum a posteriori probability 3D surface reconstruction using multiple intensity images directly*, SPIE Vol. 1260 Sensing and Reconstruction of Three-Dimensional Objects and Scenes (1990), pp. 36-48.

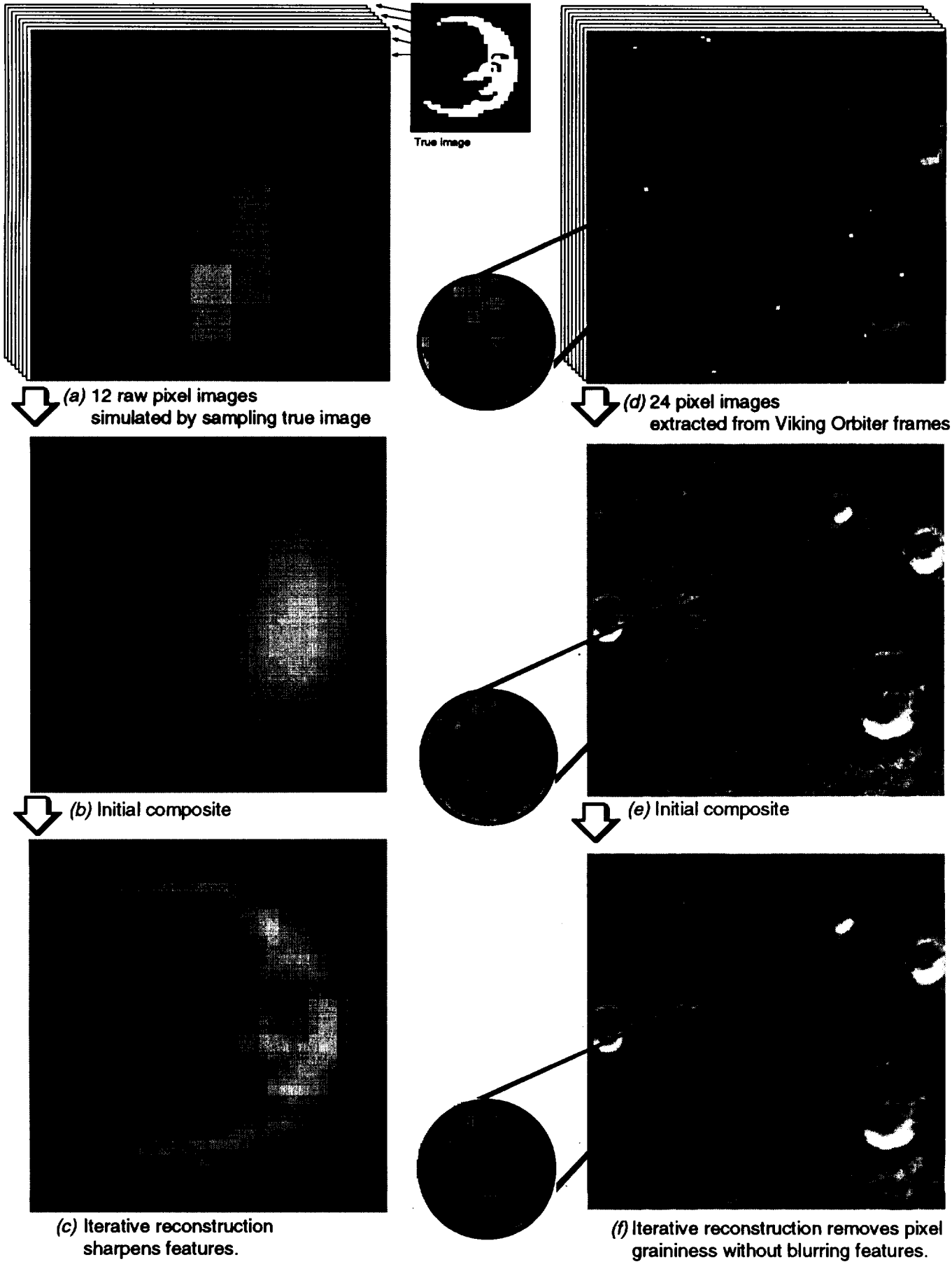

Figure 2: Surface Reconstruction